

STIC-ILL

From:
Sent:
To:
Subject:

Huynh, Phuong N.
Thursday, January 18, 2
STIC-ILL
RE: 09/270,910

11506-ES
+ C

Hi!

Please deliver the following article:

1. EMBO J. 8: 1935-38; 1989.

2. Int. Arch. Allergy Appl. Immunol. 93: 378-84; 1990.

Thanks,
"NEON" Huynh
Art Unit 1644
Tel: 308-4844
Mail: CM1 9E12

The gene coding for the major birch pollen allergen *Betv1*, is highly homologous to a pea disease resistance response gene

Heimo Breiteneder, Karin Pettenburger, Arnold Bito¹, Rudolf Valenta, Dietrich Kraft, Helmut Rumpold², Otto Scheiner and Michael Breitenbach³

Institute of General and Experimental Pathology, A-1090 Vienna,

¹Institute of Molecular Biology, Austrian Academy of Sciences, A-5020 Salzburg, and ²Institute of Clinical Chemistry and Laboratory Diagnostics and ³Institute of Microbiology and Genetics, University of Vienna, A-1090 Vienna, Austria

Communicated by G.Kreil

Pollen of the white birch (*Betula verrucosa*) is one of the main causes of Type I allergic reactions (allergic rhinoconjunctivitis, allergic bronchial asthma) in Middle and Northern Europe, North America and the USSR. Type I allergies are a major threat to public health in these countries, since 10–15% of the population suffer from these diseases. *Betv1*, an allergenic protein with an M_r of 17 kd is a constituent of the pollen of white birch and is responsible for IgE binding in more than 95% of birch pollen allergic patients. Here, we report the complete nucleotide sequence and deduced amino acid sequence of a cDNA clone coding for the major pollen allergen (*Betv1*) of white birch. It is similar to the N-terminal peptide sequences of the allergens of hazel, alder and hornbeam (close relatives) but it has no significant sequence homology to any other known allergens. However, it shows 55% sequence identity with a pea disease resistance response gene, indicating that *Betv1* may be involved in pathogen resistance of pollen.

Key words: allergy/immunology/*Betula verrucosa*/disease resistance/sequence

Introduction

Around 10–15% of the population in developed countries suffer from IgE-mediated 'atopic' diseases (Type I allergies) such as allergic rhinitis/rhinoconjunctivitis and bronchial asthma of allergic origin (Kaplan, 1985). Pollen of grasses, trees and weeds, spores of moulds as well as other airborne particles such as faeces from mites and dander from pets and other domestic animals serve as a source of allergenic proteins (Marsh *et al.*, 1986). They trigger release of inflammatory mediators (histamine, arachidonic acid metabolites and others) from mast cells and basophilic leukocytes through cross-linking of IgE bound to the cell surface via Fcε receptors (Kaplan, 1985). At present, it is not well understood what contributes to the dysregulation of the immune system leading to enhanced IgE-synthesis which is characteristic for atopic patients.

Highly purified allergens facilitate an accurate diagnosis by *in vivo* (skin-prick tests) and *in vitro* tests (radioimmunoassays or enzyme-linked immunosorbent assays). The

treatment of Type I allergies also requires exactly defined allergen solutions for hyposensitization (Creticos and Norman, 1987).

A number of attempts have been made to purify and characterize allergenic molecules by standard biochemical techniques (Mole *et al.*, 1975; Ipsen and Loewenstein, 1983; Letterman and Ohman, 1984; Chapman *et al.*, 1988; Klapper *et al.*, 1988). The purification of the allergens allowed determination of several partial N-terminal sequences, and two complete protein sequences (*AmbaV*: Mole *et al.* 1975; *AmbaIII*: Klapper *et al.*, 1988). Recently, a cDNA cloning technique was successful in the case of *Derp1*, one of the

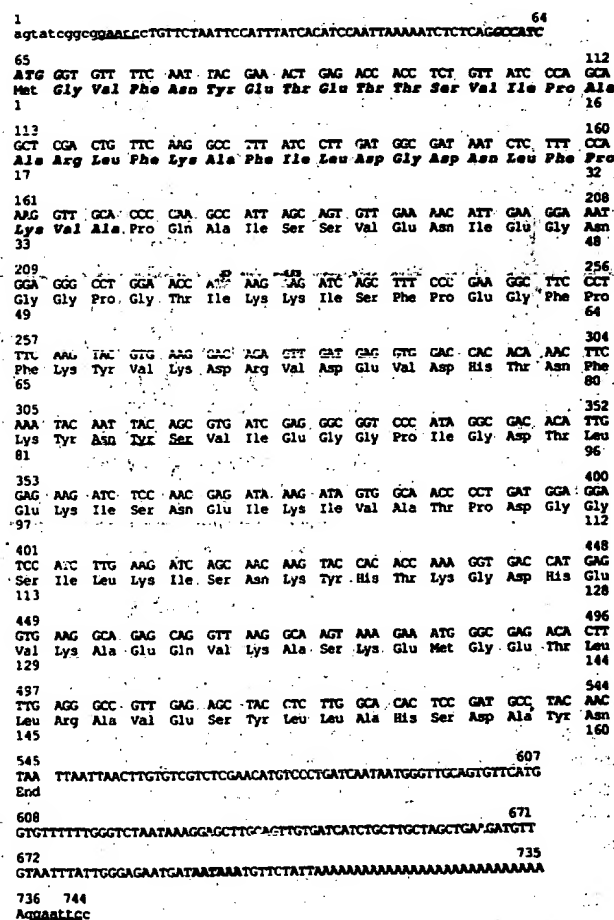


Fig. 1. Nucleotide sequence of the cDNA insert of the *Betv1* clone. Lower case letters, vector DNA including *EcoRI* linkers; bold face italic DNA sequences, the consensus surrounding the start AUG; bold face, the polyadenylation signal, underlined sequences, *EcoRI* sites. Below the DNA sequence: bold face italics, N-terminal sequence of *Betv1* as determined by Edman degradation; underlined, the single glycosylation site of *Betv1*. Numbers indicate nucleotides (above) and amino acid residues (below) of the sequence.

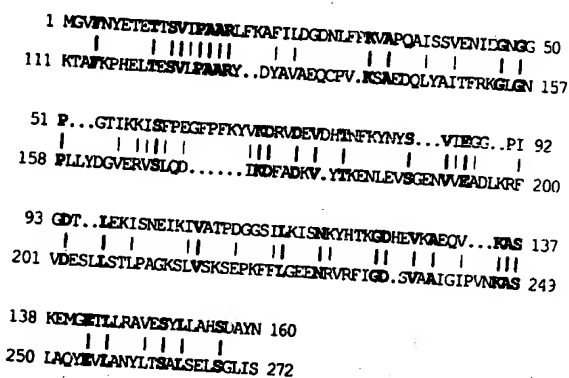


Fig. 2. Sequence comparison of *Bervl* (upper line, amino acid residues 1-160) and yeast QH2:cytochrome c oxidoreductase (lower line, amino acid residues 111-272). Identities are marked by bold face capitals and vertical bars. Conservative exchanges are marked by vertical bars. Dots indicate gaps in the sequence produced by the aligning programme. The single letter code for amino acids is used.

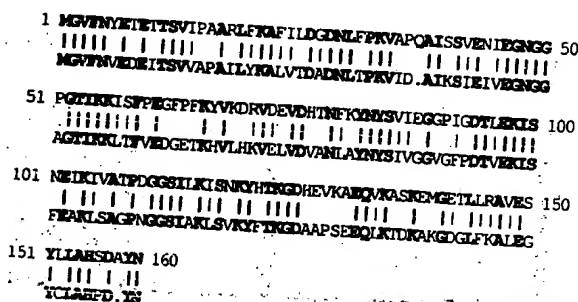


Fig. 3. Sequence comparison of *Bervl* (upper line, amino acid residues 1-160) and pea gene 149 (Fristensky et al., 1988) (lower line, amino acid residues 1-158). Identities and conservative exchanges are marked as in Figure 2.

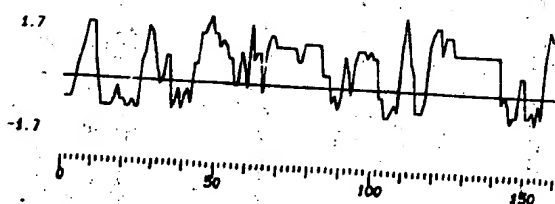


Fig. 4. Computer aided prediction of antigenic sites within the protein sequence of *Bervl*. The antigenic index according to Wolf et al. (1988) is plotted against amino acid residue number. Peaks reaching an antigenic index of ~1.7 represent possible epitopes.

major allergens of the house dust mite *Dermatophagoides pteronyssinus* (Chua et al., 1988a).

In this paper, the complete amino acid sequence of *Bervl*, the single major allergen of pollen from white birch (*Betula verrucosa*) is reported. The sequence was derived from the nucleotide sequence of a cDNA clone obtained from mature pollen mRNA and is in complete agreement with a partial N-terminal protein sequence of *Bervl* as determined by automated solid phase Edman degradation; this sequence is, to the best of our knowledge, the first published complete cDNA sequence of any allergen. Knowledge of the complete sequence of the most important allergen of tree pollen is a

prerequisite for progress in diagnosis and therapy of Type I allergic disease, as will be discussed. It also enables us, by sequence comparison, to learn about biological functions of the allergen. As it turns out in this case, *Bervl* is highly homologous to a plant disease resistance response gene of pea and may be involved in disease resistance of birch pollen.

Results

Sequence of the cDNA clone

The sequence of the cDNA insert (Figure 1) of the *Bervl* clone contains an open reading frame (ORF) of 161 codons. The deduced amino acid sequence is identical with an N-terminal protein sequence of 34 amino acid (Figure 1) as determined by solid phase Edman sequencing of *Bervl* (data kindly provided by Dr I. Maurer-Fogy). The deduced M_r of *Bervl* is 17.420 kD. The coding sequence shows a single consensus site for N-glycosylation at codon 83. The cDNA contains non-translated leader and trailer sequence at its 5'- and 3'-ends (Figure 1). The ORFs of *lacZ* and *Bervl* are joined in-frame by the linker and leader sequences enabling the fusion protein to be expressed. The six bases preceding and the base following the start AUG of *Bervl* are in agreement with consensus sequences of eukaryotic protein coding genes (Lütke et al., 1987). Our sequence also contains a long trailer followed by a canonical AATAA polyadenylation signal (Birnstiel et al., 1985), and by the poly(A)-tail.

Sequence comparisons

Comparing our sequence of *Bervl* with other known allergen sequences obtained by Edman degradation shows a very significant N-terminal homology with partial peptide sequences of the corresponding allergens of alder, hazel and hornbeam, closely related tree species (Borch et al., 1987). On the other hand, short N-terminal amino acid sequences of the allergens of dust mite, *DerfI* (Heymann et al., 1986) and cat dander, *FeldI* (Chapman et al., 1988) and two complete amino acid sequences of ragweed pollen allergens, *AmbaIII* (Klapper et al., 1988) and *AmbaV* (Mole et al., 1975), determined by Edman degradation, showed no significant homology. Furthermore, a cDNA sequence of the house dust mite allergen, *DerpI* (Chua et al., 1988a) and a partial cDNA sequence of *DerpII* (Chua et al., 1988b) showed no homology to *Bervl*. Comparing the *Bervl* sequence with all known sequences in the EMBL protein sequence data bank revealed homologies with two known protein sequences. A sequence identity of 25% (38% sequence homology including conservative exchanges) was found with the precursor of subunit II of the QH2:cytochrome c oxidoreductase (EC 1.10.2.2) of *Saccharomyces cerevisiae* (Oudshoorn et al., 1987). The homology was found over the whole length of the *Bervl* protein sequence, but the yeast protein had both N-terminal and C-terminal extensions not homologous to *Bervl* (Figure 2). A much stronger similarity was found between *Bervl* and a gene (149) involved in pea disease resistance response (Fristensky et al., 1988). The proteins encoded by the two genes are of nearly identical size (160 and 158 amino acids, respectively) and show 55% sequence identity (70% sequence similarity including conservative exchanges) (Figure 3).

Structure predictions

A computer aided prediction of antigenic sites with the protein sequence of *Betv1* showed possible epitopes for antibody binding near residues 10, 30, 50, 110, 120 and 155 (Figure 4). The predictions based on two methods (Garnier *et al.*, 1978; Wolf *et al.*, 1988) are in reasonable agreement.

Discussion

The first 35 amino acids of the sequence of *Betv1* as obtained by us and others (Ipsen and Hansen, 1987) are identical with the respective protein sequence deduced from the complete cDNA sequence (Figure 1). There is indirect but suggestive evidence that the rest of the ORF coding for *Betv1* is also correct: The total length of the ORF (159 amino acids) and its calculated M_r of 17.420 kd correlates very well with the observed M_r as determined by SDS-PAGE and gel chromatography (Ipsen and Loewenstein, 1983; Jarolim *et al.*, 1989b). Furthermore, most of the C-terminal half of the *Betv1* sequence has been determined by sequencing both DNA strands.

Comparing the six bases preceding and the base following the start AUG of *Betv1* with consensus sequences of eukaryotic protein coding genes (Lütcke *et al.*, 1987) shows that our sequence is nearly identical (only one mismatch out of 10 bp) with the animal consensus sequence and surprisingly bears little resemblance to the plant consensus. The presence of the start AUG codon in its canonical surroundings shows that the clone is complete in its 5'-end and that the N-terminus of the protein is not subject to post-translational processing.

The sequence of *Betv1* is homologous with other known partial sequences from the N-terminus (~40 amino acids) of the allergens of alder, hazel and hornbeam (Borch *et al.*, 1987). This reflects the taxonomic relationship between these tree species (order *Fagales*). The predicted first two epitopes of *Betv1* are in the region of strong homology between birch and the N-terminal hazel and alder pollen allergen sequences. It will be of particular interest to compare the results of the predictions with the results of the epitope mapping experiments presently in progress.

Betv1 shows weak but significant homology to the protein sequence of yeast QH2:cytochrome *c* oxidoreductase, a component of the inner mitochondrial membrane (Figure 2). The homology of *Betv1* to a pea disease resistance response gene (*I49*) is much more pronounced, although pea (*Fabaceae*) is quite unrelated to birch (*Betulaceae*) (Figure 3). Both, *Betv1* and *I49* are members of families of structurally closely related genes. It is not unlikely, in our view, that *Betv1* fulfills a function similar to gene *I49* of pea which is induced in pea pods upon contact with the plant pathogens *Fusarium solani* f.sp. *pisi* (a fungus) and *Pseudomonas syringae* pv. *pisi* (a bacterium), respectively (Fristensky *et al.*, 1988). However, it is unknown at present whether the *I49* gene product is directly active in plant defence against *Fusarium* or *Pseudomonas* (Lamb *et al.*, 1989). Experiments are in progress in our laboratory to investigate a possible antifungal or antibacterial activity of purified *Betv1* protein. A pathogen defense function of *Betv1* would be in accordance with our recent results showing that *Betv1* and its mRNA are found in several somatic tissues of white birch. *Betv1* is present in pollen and also in low concentration in leaves. The mRNA for *Betv1* was not only found in pollen and

leaves, but also in male and female inflorescences and phytohormone-dependent plant tissue culture (K. Pettenburger *et al.*, in preparation).

The coding sequence shows a single consensus site for *N*-glycosylation at codon 83. We are presently trying to chemically analyse the sugars bound to the *Betv1* protein purified by standard methods. In fact, however, we know that the sugar residues do not constitute epitopes recognized by patients' IgE, since *Betv1* synthesized *in vitro* from pollen mRNA is fully reactive with IgE from patients allergic to birch pollen (Breiteneder *et al.*, 1988). In addition, as described in this paper, the β -galactosidase-*Betv1* fusion protein (that is also not glycosylated) synthesized in *Escherichia coli* is also reactive with patients IgE.

As concluded from genomic Southern blots, *Betv1* may be encoded by a small gene family (data not shown). This observation is in accordance with data from immunoblotting experiments after two-dimensional PAGE of *Betv1* purified by immunoaffinity chromatography. *Betv1* was resolved into ~10 IgE-binding isoallergens (Jarolim *et al.*, 1986b) that obviously all shared the same N-terminal amino acid sequence since Edman degradation of the mixture of the isoforms showed single homogenous peaks for each determined amino acid.

Cloning of genes coding for allergens and determination of the gene sequences seems a prerequisite for answering basic questions regarding pathogenesis of Type I allergies. Knowledge of these sequences seems necessary for investigating the major theoretical problems of characterizing the epitopes recognized by T and B lymphocytes, respectively, which are responsible for the enhanced IgE response in atopic patients.

Epitope mapping can be done by a number of techniques, but preferentially by testing the reactivity of short recombinant or synthetic peptides, derived from the sequence.

Exact testing of patients' sera is necessary for a reliable diagnosis and successful treatment (hyposensitization) of the disease. Some cases of unsuccessful immunotherapy may be due to the absence of the specific allergens to which the patient is most responsive from solutions used for testing and hyposensitization. This can best be helped by producing standardized allergens by recombinant DNA techniques.

Knowing a short but fully reactive monovalent peptide sequence out of the total sequence of the allergen might lead to a radically new form of therapy. Such a monovalent peptide might competitively block IgE bound to mast cells' IgE-Fcε receptors and, thus, prevent their cross-linking and subsequent mediator release.

Materials and methods

Poly(A)⁺ mRNA isolation and cDNA cloning

Total pollen RNA was extracted from 500 mg mature birch pollen (Allergon AB, Engelholm, Sweden) (Mascarenhas *et al.*, 1984). Standard scanning electron microscopic techniques showed that the pollen was of the species *B. verrucosa* and free of contaminations. Poly(A)⁺ RNA was isolated by binding to oligo(dT)-cellulose (Sigma, St Louis, MO) (Ausubel *et al.*, 1987a). A cDNA-synthesis- and a cDNA-cloning-kit (Amersham Int., Little Chalfont, England) were used according to the manufacturer's recommendations.

Screening of the cDNA expression library in bacteriophage λgt11

Plaques (0.5×10^6) were immunologically screened for the expression of *Betv1* using the serum of a birch pollen allergic patient. The patient was selected for case history, positive skin-prick test and positive radioallergen sorbent test and had not undergone any hyposensitization treatment. Immunoblotting ex-

periments showed that IgE of this serum reacted exclusively with the *Berl* of an aqueous extract of birch pollen (Jarolim et al., 1989a). Plaques were lifted onto nitrocellulose filters (Schleicher and Schuell, Dassel, FRG) and the filters washed twice in 30 ml of buffer (50 mM Na-phosphate, pH 7.5, 0.5% v/v Tween 20, 0.5% bovine serum albumin (BSA), 0.05% Na₂S₂O₅ w/v) for 5 min to remove pieces of agar and then again incubated in buffer for 30 min to block remaining binding sites on the surface of the filters. The filters were then incubated at 4°C overnight in buffer containing 10% patient's serum. The filters were washed 3 times with 30 ml of buffer and incubated with ¹²⁵I-labelled anti-human IgE (Pharmacia Diagnostics AB, Uppsala, Sweden) diluted 10-fold in buffer (25 ml per filter). The incubation was performed overnight at room temperature under slight agitation. The filters were again washed 3 times, dried, and positive clones were detected by autoradiography. After recloning, two clones remained positive; they seemed identical as judged by restriction analysis.

Phage DNA preparation

DNA was prepared from the phages of the plaque-purified clones following the liquid lysate method (Ausubel et al., 1987).

Construction of lysogens and immunoblot analysis of the *Berl*- β -galactosidase fusion protein

Lysogenic bacterial cells infected with the recombinant λ gt11 clone were grown under appropriate conditions stimulating fusion protein synthesis (Glover, 1985). The cells were then harvested at room temperature and resuspended in 1/30 of the original volume of phosphate buffer (50 mM Na-phosphate, pH 7.5), frozen in liquid nitrogen and thawed at 37°C. The lysed bacterial cells were lyophilized and crude bacterial proteins (500 μ g) from each induced lysogen were separated by 12% SDS-PAGE and electroblotted onto nitrocellulose (Towbin et al., 1979). The blots were washed, incubated in buffer and then incubated in the selected patient's serum mentioned above. IgE-binding to the β -galactosidase-*Berl* fusion protein was detected by ¹²⁵I-labelled anti-human IgE. Methods for washing, blocking and antibody incubation were the same as described for the screening of the λ gt11 cDNA library.

Subcloning of the λ gt11 inserts

Since the left *Eco*RI site was destroyed during cloning (as seen in Figure 1), 1 μ g DNA from each clone was cut with *Kpn*I and *Sac*I releasing a 2.8 kb fragment. This fragment was ligated into the polycloning site of the plasmid Bluescript SK⁺ (Stratagene, San Diego, CA) and used to transform *E. coli* XL1-Blue (Stratagene). This subclone was designated pBV1.

Sequencing of the cDNA insert

(i) Double strand sequencing of pBV1 was performed according to the modified chain terminating method (Zagursky et al., 1985) using two commercially available λ gt11 sequencing primers (Clontech, Palo Alto, CA) and the enzyme, sequenase (Stratagene). (ii) As the sequence could not be read unambiguously in the vicinity of the singular *Bgl*II site of the cDNA insert, plasmid pBV1 was digested with *Bgl*II and the 3'-overhangs were filled in with [³²P]dATP using reverse transcriptase. The labelled DNA was redigested with *Sac*I, thus yielding two labelled fragments, which were sequenced according to the protocol of Maxam and Gilbert, 1970. (iii) Plasmid pBV1 was digested with *Eco*RI-*Bgl*II and with *Bam*HI-*Hinc*II. (Enzyme *Hinc*II cuts at the destroyed *Eco*RI site.) The resulting fragments were subcloned into the plasmids Bluescript SK⁺ and Bluescript SK⁻. Helper phage R408 (Stratagene) was used to isolate the single-stranded forms of the subcloned fragments. Sequencing of the single-stranded DNAs was performed with the Bluescript T3 and T7 primers using the enzyme, sequenase and the modified chain terminating method (Zagursky et al., 1985).

Genomic Southern blots

Genomic DNA was isolated from birch leaves (Dellaporta et al., 1983), cut with *Bam*HI, *Eco*RV and *Hind*III, and separated electrophoretically in 0.7% agarose. The DNA fragments were blotted to nitrocellulose and probed with a ³²P-labelled *Eco*RI-*Hinc*II-fragment of pBV1. For each enzyme at least four fragments hybridizing with the probe were visualized by autoradiography (Southern, 1978).

Acknowledgements

Our thanks are due to I. Maurer-Fogy for protein sequencing, M. Tejkl for supplying purified *Berl*, I. Lubenik for secretarial assistance, and to M. Duchêne and M. Grinner for their help with computer work. Financial support by Mayerhofer and Co., Linz, Austria and by the 'Forschungsfonds für die Gewerbliche Wirtschaft' (grant number 3/6197) is gratefully acknowledged.

References

- Ausubel, F.M. (ed.) (1987a) In *Current Protocols in Molecular Biology*. Greene Publishing Associates and Wiley-Interscience: John Wiley and Sons, New York, Unit 4.5.
- Ausubel, F.M. (ed.) (1987b) In *Current Protocols in Molecular Biology*. Greene Publishing Associates and Wiley-Interscience: John Wiley and Sons, New York, Unit 1.13.5.
- Birnsteil, M.L., Busslinger, M. and Strub, K. (1985) *Cell*, **41**, 349-359.
- Borch, S.M., Sletten, K. and Hagen, A.M. (1987) In 16th Nordic Congress on Allergy, Tromsø, Abstract VI-19, p. 65, June.
- Breiteneder, H., Hassfeld, W., Pettenburger, K., Jarolim, E., Breitenbach, M., Rumpold, H., Kraft, D. and Scheiner, O. (1988) *Int. Archs. Allergy Appl. Immunol.*, **87**, 19-24.
- Chapman, M.D., Aalberse, R.C., Brown, M.J. and Platts-Mills, T.A.E. (1988) *J. Immunol.*, **140**, 812-813.
- Chua, K.Y., Stewart, G.A., Thomas, W.R., Simpson, R.J., Dilworth, R.J., Plozza, T.M. and Turner, K.J. (1988a) *J. Exp. Med.*, **167**, 175-182.
- Chua, K.Y., Boyle, C.R. and Dilworth, K.J. (1988b) *N. Engl. Reg. Allergy Proc.*, **9**, 415 (Abstract).
- Creticos, B.S. and Norman, P.S. (1987) In Lessof, M.H., Lee, T.H. and Kemeny, (eds), *Allergy: An International Textbook*. John Wiley and Sons, Chichester, pp. 617-630.
- Dellaporta, S.L., Wood, J. and Hicks, J.B. (1983) *Plant Mol. Biol. Rep.*, **1**, 19-21.
- Fristensky, B., Horovitz, D. and Hadwiger, L.A. (1988) *Plant Mol. Biol.*, **11**, 713-715.
- Garnier, J., Osguthorpe, O.J. and Robson, B. (1978) *J. Mol. Biol.*, **121**, 97-120.
- Glover, D.M. (ed.) (1985) In *DNA Cloning. A Practical Approach*. IRL Press, Oxford, Vol. 1.
- Heymann, P.W., Chapman, M.D. and Platts-Mills, T.A.E. (1986) *J. Immunol.*, **137**, 2841-2847.
- Ipsen, H. and Hansen, V.C. (1987) In 16th Nordic Congress on Allergy, Tromsø, Abstract VI-II, p. 61, June.
- Ipsen, H. and Loewenstein, H. (1983) *J. Allergy Clin. Immunol.*, **72**, 150-159.
- Jarolim, E., Rumpold, H., Endler, A., Ebner, H., Scheiner, O. and Kraft, D. (1989a) *Allergy*, **44**, 876-885.
- Jarolim, E., Rumpold, H., Endler, A., Schlerka, G., Ebner, H., Scheiner, O. and Kraft, D. (1989b) *Int. Archs. Allergy Appl. Immunol.*, **88**, 180-182.
- Kaplan, A.P. (ed.) (1985) *Allergy*. Churchill Livingstone, New York.
- Klapper, D.G., Goodfriend, L. and Capra, J.D. (1988) *Biochemistry*, **29**, 5729-5734.
- Lamb, C.J., Lawton, M.A., Dron, M. and Dixon, R.A. (1989) *Cell*, **56**, 215-224.
- Letterman, K. and Ohman, J.L. (1984) *J. Allergy Clin. Immunol.*, **74**, 147-153.
- Lütcke, H.A., Chow, K.C., Mickel, F.S., Moss, K.A., Kern, H.F. and Scheele, G.A. (1987) *EMBO J.*, **6**, 43-48.
- Mascarenhas, N.T., Bashe, D., Eisenberg, A., Willing, R.P., Xiao, C.-M. and Mascarenhas, J.P. (1984) *Theor. Appl. Genet.*, **68**, 323-326.
- Marsh, D.G., Goodfriend, L., King, T.P., Loewenstein, H. and Platts-Mills, T.A.E. (1986) *Bull. WHO*, **64**, 767-770.
- Maxam, A.M. and Gilbert, W. (1980) *Methods Enzymol.*, **65**, 499-560.
- Mole, L.E., Goodfriend, L., Lapkoff, C.B., Kehoe, J.M. and Capra, J.D. (1975) *Biochemistry*, **14**, 1216-1220.
- Oudshoorn, P., Van Steeg, H., Swinkels, B.W., Schoppink, P. and Grivell, L. (1987) *Eur. J. Biochem.*, **163**, 97-103.
- Southern, E. (1978) *J. Mol. Biol.*, **98**, 503-505.
- Towbin, H., Staehelin, Th. and Gordon, J. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 4350-4354.
- Wolf, H., Modrow, S., Moth, M., Jameson, B.A., Hermann, G. and Förtsch, B. (1985) *CABIOS*, **4**, 187-191.
- Zagursky, R., Baumeister, N., Lomax, N. and Bergman, M. (1985) *Gene Anal. Technol.*, **2**, 89-94.

Received on February 27, 1989; revised on March 30, 1989